



Predicting Lymph Node Metastasis in T1 Colorectal Cancer Patients Using Interpretable Machine Learning Models: A Multicenter Retrospective Study

Qingyang Fang¹, Xinyang He^{2*}

Department of General Surgery, The Affiliated Provincial Hospital of Anhui Medical University, Hefei 23000, China

Corresponding Author Email: hxy2333@126.com

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/rces.120101>

ABSTRACT

Received: 13 January 2025

Revised: 25 February 2025

Accepted: 4 March 2025

Available online: 31 March 2025

Keywords:

early colorectal cancer, lymph node metastasis, machine learning

Whether lymph node metastasis (LNM) is present is crucial for treatment decisions in T1 colorectal cancer (T1 CRC). This study developed predictive models using data from 1,205 patients across seven Chinese medical centers. We evaluated 29 machine learning algorithms and identified CatBoost as the top performer (AUC: 86%, accuracy: 96%). SHAP analysis revealed key predictors of LNM risk, including lymphovascular invasion, age, tumor size, invasion depth, and total lymph node count. Less influential features included perineural invasion and tumor location. The study highlights the importance of retrieving more lymph nodes during surgery to improve staging accuracy. A user-friendly online tool was developed to support clinical decision-making.

1. INTRODUCTION

Globally, CRC is recognized as the third most frequently diagnosed malignancy and ranks second in terms of cancer-related mortality [1]. With widespread CRC screening, diagnosed cases of T1 CRC have increased [2, 3]. In China, CRC has become the second most common malignant tumor (29.51 per 100,000) and the fourth primary cause of cancer-associated fatalities (14.14 per 100,000) [4]. LNM, defined as cancer cells found in lymph nodes surrounding the primary tumor, indicates tumor spread [5]. Clinically, the presence of LNM dictates treatment: patients without LNM may achieve curative results through endoscopic resection, whereas those with LNM require radical surgery [6]. Currently, LNM evaluation mainly relies on MRI, CT, and biopsy [7-9]. However, these methods fail to detect LNM in 8-12% of patients due to imaging limitations [10], and their high cost and long duration limit widespread use. Hence, there is an urgent need for more efficient, imaging-independent detection methods. Compared to traditional logistic regression (LR), recent advancements in statistical theory and machine learning have led to predictive models with superior performance, enhancing prediction of LNM in cancers such as gastric, thyroid, and breast [11-13].

The history of machine learning dates to 1957, when Rosenblatt [14] proposed the “Perceptron”, recognized as the earliest machine learning model. Simultaneously, LR, developed between the late 19th and early 20th centuries, gained prominence for binary classification tasks [15]. LR remains a common baseline method, though it is often outperformed by more complex models like Support Vector Machines (SVMs) and Neural Networks (NNs). Linear models such as logistic and Least Absolute Shrinkage and Selection Operator (LASSO) regression have limitations. Hartwig et al.

[16] employed a LASSO model using data from 35,812 Danish patients, achieving a relatively low AUC (0.64), indicating poor model selection. Wang et al. [17] used LR on 825 T1-stage CRC patients, achieving an AUC of 0.793, yet faced overfitting concerns due to limited data. Fujino et al. [18] assessed LNM risks with LR on 934 Japanese patients, yielding an AUC of 0.786 (training) and 0.721 (validation), but lacked robustness due to the absence of cross-validation. Similarly, Niu and Cao [19] achieved an AUC of only 0.708 from a large SEER database without cross-validation, limiting its generalizability.

SVM is a machine learning method using optimization to avoid the “curse of dimensionality” and “overfitting” by finding an optimal hyperplane that maximally separates classes. Ichimasa et al. [20] analyzed data from 690 T1-stage CRC patients (590 training, 100 validation) between 2001 and 2016. Their model incorporated 45 clinicopathological factors—such as age, Body Mass Index (BMI), tumor size, location, LVI, tumor markers (CEA, CA19-9), and biochemical indicators—to predict LNM risk. ROC analysis showed good accuracy (AUC=0.821). However, small sample size, low LNM prevalence (55 positive cases in training, 9 in validation), data imbalance, and inclusion of clinically ambiguous features limited model reliability and practical applicability.

The training mechanism of NNs comprises two essential phases: forward propagation of signals and back-propagation of errors. During the forward pass, input data traverse through multiple hidden layers before reaching the output layer. If actual outputs differ from expected, error signals propagate backward for weight adjustments. NNs evolved into models like TabNet, an NN combining attention mechanisms and tree-based feature selection. TabNet’s main steps include (a) sparse feature selection; (b) sequential multi-step structures for

incremental decision contributions; (c) nonlinear processing to enhance learning; and (d) ensemble processing through multiple steps. Song et al. [21] developed a deep learning model leveraging attention mechanisms to analyze whole-slide images (WSIs) for predicting LNM in T1 CRC, attaining AUC values ranging from 0.781 to 0.824. Despite promising results, limitations include single-center WSI data, small samples, potential preparation biases, complexity in interpreting high-dimensional data, and challenging feature associations, complicating model interpretability.

Decision Trees (DTs) are flowchart-like structures with nodes representing attribute tests, branches representing outcomes, and leaves representing class distributions. AdaBoost sequentially trains a series of weak learners and aggregates them into a composite strong model. Random Forest (RF) integrates multiple DTs via bagging, while Gradient Boosted Decision Trees (GBDTs) iteratively aggregate weak learners for improved predictions. Takamatsu et al. [22] proposed a hybrid approach, combining Convolutional Neural Networks (CNNs) with RF classifiers to predict LNM from histopathological WSIs in T1 CRC cases. Among 783 cases divided into training (548) and validation (235), AUC values were 0.971 (training) and 0.760 (validation). However, training the CNN on roughly 500 cases led to an AUC drop of 21%, indicating significant overfitting and limited reliability.

XGBoost, proposed by Chen [23], is a powerful tree-based ensemble learning algorithm widely adopted for classification and regression problems. Unlike RF and GBDT, XGBoost applies first- and second-order derivatives via a second-order Taylor expansion of the loss function, introducing L1 and L2 regularization and column sampling to reduce overfitting. XGBoost represents trees using leaf weights, directly expressing the loss function's extremum through derivatives [24]. The "gain" metric, reflecting the relative contribution of features, determines optimal split points. Ahn et al. [25] applied XGBoost to data from 26,733 T1-stage CRC patients (SEER database), using eight prognostic variables and achieving a modest AUC of 0.659 (sensitivity 0.242, accuracy 0.604, F1 score 0.352). Despite the large sample and relevant variables, the low performance suggests inadequate parameter tuning, limiting the model's reliability.

The Light GBM model, a highly optimized implementation of the GBDT algorithm, is designed to handle large-scale data with superior computational efficiency. It supports parallel learning, operates at high speed, consumes less memory, and delivers improved predictive accuracy. In a study conducted by Piao et al. [26], a LightGBM-driven model was constructed using data from 651 patients diagnosed with T1-stage CRC who underwent radical resection across six medical institutions in Ningbo between 2016 and 2022. The model demonstrated excellent discriminative ability, with an AUC of 0.960. Despite this high performance, the small sample size raises concerns about overfitting. Additionally, reliance on pathologist-dependent features, such as submucosal invasion area, limits the model's standardization and broader applicability.

CatBoost, a gradient boosting algorithm released as an open-source project by the Russian tech company Yandex in 2017, represents another advancement in the GBDT family. While XGBoost is widely used and LightGBM improves computational efficiency, CatBoost claims superior accuracy [27]. It employs symmetric DTs (oblivious trees) as base

learners, featuring fewer parameters and built-in support for categorical variables. This enables efficient handling of categorical data—reflected in its name, combining "Categorical" and "Boosting." Furthermore, CatBoost mitigates issues such as gradient bias and prediction shift, thereby reducing the likelihood of overfitting and enhancing both accuracy and model generalizability.

Machine learning methods are advantageous in capturing higher-order nonlinear interactions among predictive factors, often leading to more stable and robust predictions. However, their inherent "black box" nature makes it difficult to explain the rationale behind specific predictions, limiting their acceptance in clinical decision-making. SHAP addresses this challenge by precisely quantifying the contribution and influence of each input feature on the model's output [28]. While numerous machine learning models have shown high performance, a recurring shortcoming in previous studies is the absence of transparent and intuitive feature attribution. Without clear attribution, clinicians may struggle to identify key features, limiting the model's practical utility in guiding treatment decisions.

2. RESULTS

2.1 Comparison of model performance

A total of 29 distinct machine learning algorithms were constructed to estimate the likelihood of LNM in individuals diagnosed with T1 CRC. The evaluation metrics for each model are summarized in Table 1. To assess and compare predictive capabilities, multiple performance indicators were employed, including accuracy (ACC), precision, recall, F1 score, and AUC.

In this study, except for CatBoost and Random Forest, the other 27 models showed poor performance in distinguishing LNM, with AUC values below 0.80 and notably low precision and F1 scores. This highlights their limited ability to correctly identify positive cases, often resulting in high false-positive rates. Although Perceptron and SVM achieved relatively high recall, their low precision indicates a lack of reliable positive predictions. Models such as MultinomialNB, BernoulliNB, and LinearSVC failed completely, with all key metrics at zero.

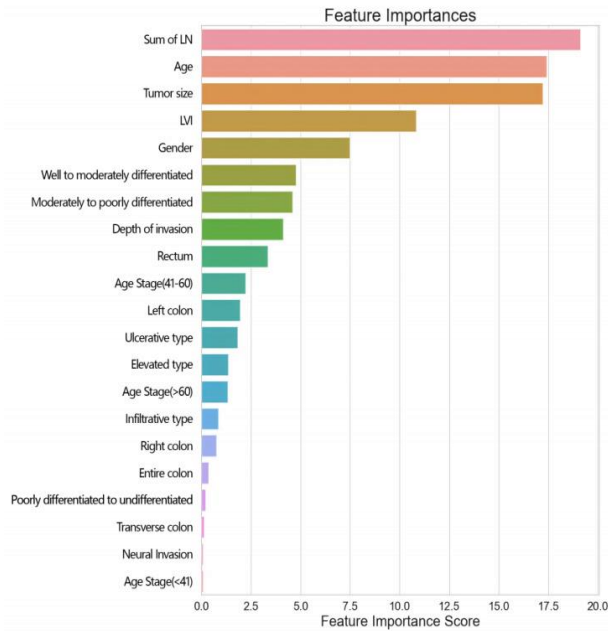
Comparatively, CatBoost significantly outperformed Random Forest. While both achieved an AUC of 0.86, CatBoost demonstrated much higher precision (0.50 vs. 0.00) and F1 score (0.29 vs. 0.00), indicating better accuracy and robustness in identifying LNM. Despite sharing the same overall accuracy (0.96), CatBoost provided more meaningful and clinically applicable predictions, whereas Random Forest's lack of precision rendered it less effective. Overall, CatBoost proved to be the more reliable and practical model for LNM classification.

2.2 Feature importance and model interpretation

The CatBoost model comes with a powerful interpretability function that can automatically generate feature importance plots (Figure 1). The features are ranked in order of importance, with the length representing the degree of importance—the longer the bar, the more significant the feature.

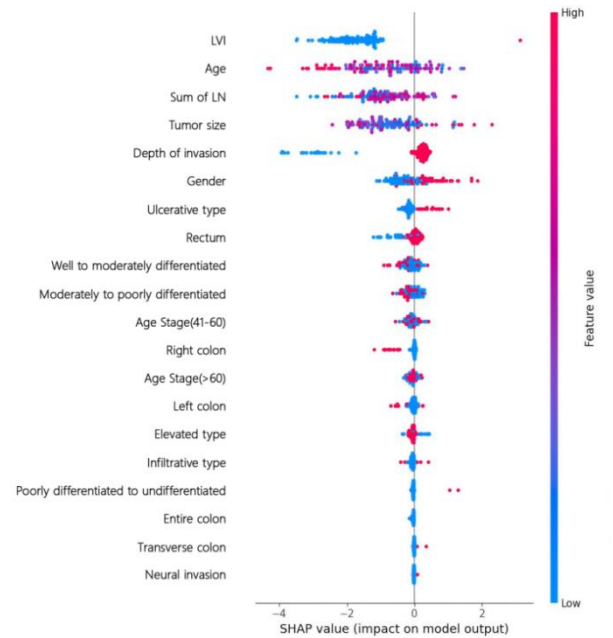
Table 1. Model evaluation results

Model	ACC	Precision	Recall	F1	AUC
Random Forest	0.96	0.0	0.0	0.0	0.86
CatBoost	0.96	0.5	0.2	0.29	0.86
AdaBoost	0.96	0.5	0.4	0.44	0.79
Bagging	0.97	1.0	0.2	0.33	0.79
XGBoost	0.94	0.25	0.2	0.22	0.79
Gradient Boosting	0.96	0.5	0.2	0.29	0.75
Extra Trees	0.95	0.0	0.0	0.0	0.73
K Neighbors	0.94	0.0	0.0	0.0	0.71
LightGBM	0.95	0.33	0.2	0.25	0.71
MLP Classifier	0.97	1.0	0.2	0.33	0.7
Decision Tree	0.94	0.33	0.4	0.36	0.68
Gaussian Process	0.95	0.0	0.0	0.0	0.68
MultinomialNB	0.95	0.0	0.0	0.0	0.64
ComplementNB	0.69	0.08	0.6	0.14	0.64
Perceptron	0.04	0.04	1.0	0.08	0.64
SGD Classifier	0.96	0.5	0.2	0.29	0.63
Linear-Discriminant-Analysis	0.97	1.0	0.2	0.33	0.62
Passive Aggressive	0.96	0.0	0.0	0.0	0.62
BernoulliNB	0.95	0.0	0.0	0.0	0.62
LCE Model	0.97	1.0	0.2	0.33	0.62
GaussianNB	0.9	0.11	0.2	0.14	0.61
Quadratic-Discriminant-Analysis	0.62	0.06	0.6	0.12	0.61
RidgeClassifier	0.88	0.08	0.2	0.12	0.59
Logistic Regression	0.84	0.06	0.2	0.1	0.59
LinearSVC	0.96	0.0	0.0	0.0	0.58
TabNet	0.97	1.0	0.2	0.33	0.5
Nearest Centroid	0.56	0.04	0.4	0.07	0.48
Extra Tree	0.9	0.0	0.0	0.0	0.47
SVM	0.56	0.06	0.6	0.1	0.31

**Figure 1.** The feature importance ranking chart illustrating the relative importance of different clinical features in the predictive model

To determine the relative significance of each feature, SHAP values were derived from the output of the CatBoost predictive model (Figure 2). This analytical procedure involves the stepwise incorporation of variables, commencing with the most influential and progressively including features of lesser importance according to their ranking. In the resulting plot, each dot corresponds to the SHAP value associated with a particular feature for an individual patient. The horizontal

placement of the dots reflects the direction and magnitude of the feature's effect on the prediction—points positioned to the right of the vertical axis indicate a positive contribution, whereas those on the left signify a negative impact. The color gradient represents the actual feature values, with red denoting higher values and blue indicating lower ones.

**Figure 2.** The SHAP summary plot demonstrating the influence of various clinical and pathological features on the model's output

Two patients were randomly selected for case analysis:

They were predicted to be LNM positive and negative, respectively. LVI has the largest positive contribution to the prediction (+3.13), while age (Age=61) has a significant

negative contribution (0.62) (Figure 3). LVI has the largest negative contribution to the prediction (2.67), followed by the total number of lymph nodes (1.65) and age (0.75) (Figure 4).

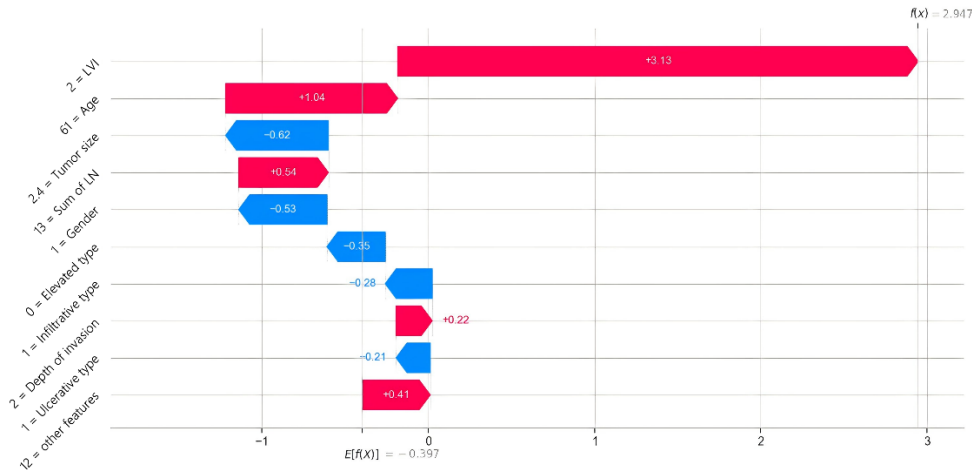


Figure 3. LNM positive

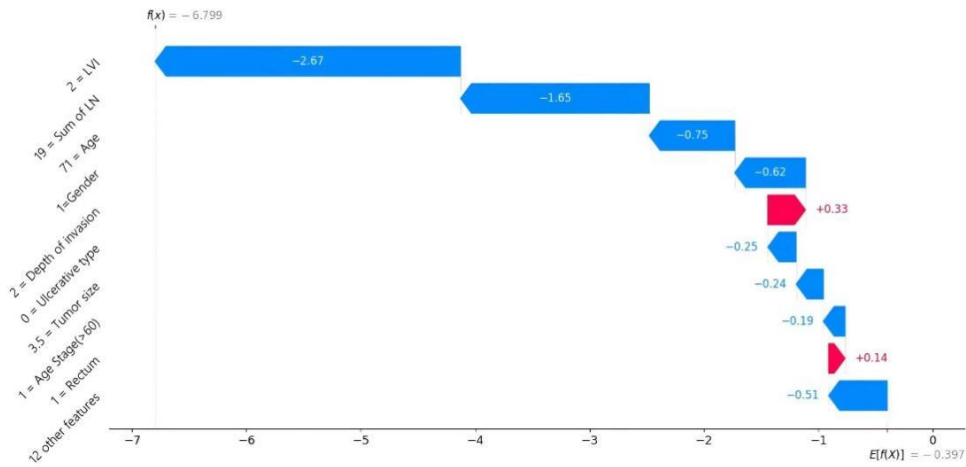


Figure 4. LNM negative

The SHAP value plot shows each feature's contribution to the model's prediction—red indicates contribution towards LNM (+), and blue towards LNM (-). $F(x)$ represents the log-odds ratio for each observation. Arrows illustrate each factor's influence: blue for decreasing LNM risk and red for increasing it. The longer the arrow, the greater the impact.

3. DISCUSSION

Accurate prediction of LNM remains a clinical challenge for patients with T1 CRC. Although the incidence of LNM is relatively low (approximately 10.9%), many patients still undergo radical surgery, which may be unnecessary in the absence of LNM [29]. Endoscopic resection has become an alternative treatment option for T1 CRC patients without LNM [30]. However, patients with T1 CRC who have concomitant LNM generally have a poorer prognosis. Therefore, establishing an accurate, reliable, and reproducible method for predicting LNM risk is crucial for optimizing preoperative

treatment strategies and reducing overtreatment [31-33]. In this study, we combined interpretable machine learning techniques with common demographic and pathological data to develop a model that predicts the LNM risk in patients with T1 CRC, providing an innovative approach for personalized treatment decisions.

T1 CRC refers to tumors confined to the mucosal or submucosal layers without reaching or penetrating the muscularis propria [7, 34]. Currently, the following five high-risk factors are generally used to guide clinical practice: tumor invasion depth, LVI, tumor differentiation grade, tumor budding, and incomplete or positive resection margins [7]. Tumor invasion depth is one of the most important high-risk factors: the deeper the invasion into the submucosa, the higher the likelihood of cancer cells invading lymph nodes. It is generally considered that tumor invasion exceeding 1,000 micrometers (1mm) is a high-risk marker for LNM [31]. LVI refers to cancer cells invading lymphatic or blood vessels, indicating that cancer cells have already spread through these vessels and may metastasize to distant sites [35]. Poorly

differentiated or undifferentiated cancer cells (including mucinous adenocarcinoma or signet ring cell carcinoma) often have higher aggressiveness and are more likely to metastasize to lymph nodes [10]. Tumor budding refers to isolated or small clusters of cancer cells formed at the tumor margin, which is a marker of tumor aggressiveness and metastatic behavior; high levels of tumor budding are significantly associated with an increased risk of LNM [36]. If pathological examination shows incomplete tumor resection margins (R1 resection) or positive margins, it implies that residual tumor cells may be present, increasing the risk of LNM [37]. If any of the above high-risk factors are found in a patient's specimen, salvage radical surgery is recommended [7]. Even so, a considerable proportion of patients will undergo unnecessary radical surgery. Researchers have reported that T1 cancer patients with high-risk factors account for 70-80% of all T1 CRC cases. Therefore, considering that only 8-12% of patients have true LNM, more than 60% of patients still receive overtreatment [10, 38]. In the guidelines for CRC treatment by the Japanese Society for Cancer of the Colon and Rectum (JSCCR), when using the predictive model in the guidelines to assess LNM in T1 CRC patients, the AUC value is about 0.588 [7]. This indicates that in practical applications, the model has certain reliability in predicting LNM in T1 CRC patients but also has significant room for improvement. Therefore, we need to develop a more accurate and feasible algorithmic model to assist clinicians in formulating surgical plans.

Our study shows that the CatBoost model achieves the highest AUC, F1 Score, Recall, Precision, and Accuracy, demonstrating the best predictive performance among the 29 models compared. Traditional machine learning models, such as artificial Neural Networks (ANNs), SVMs, and RFs, face a fundamental issue known as the "black box problem" [39]. While these models can calculate metastasis probabilities based on patient input data, their internal decision-making processes are difficult to interpret. This makes it challenging for clinicians to clearly understand which demographic characteristics or pathological report details play key roles in the prediction, and they cannot effectively utilize this information to decide on additional treatments. Therefore, the clinical application of these models is somewhat limited, making it difficult to fully meet the medical field's demand for transparency and interpretability in decision-making.

The CatBoost model effectively addresses this problem. It relies on changes in the internal loss function and pre-diction shifts within the model to evaluate feature importance, offering high computational efficiency and being particularly suitable for tree-based models. However, it primarily provides global interpretations and cannot deeply explain the feature contributions for each individual prediction [27]. In contrast, SHAP, based on Shapley values from cooperative game theory, provides consistency and local interpretability by considering the marginal contributions of features. It is applicable to various model types but has higher computational complexity when handling a large number of features [28]. Therefore, CatBoost is more suitable for scenarios requiring rapid global feature importance, while SHAP offers more comprehensive global and local explanations [40]. To reflect the contributions of all factors in LNM, we used machine learning combined with the SHAP method to evaluate the risk of LNM in T1 CRC patients. This method displays a list of important features, ranked from most to least important (from top to bottom). Consequently, the important features identified by CatBoost were further plotted.

The results showed that the model's predictions are significantly influenced by features such as LVI, age, total number of lymph nodes, tumor size, and depth of tumor invasion, while features like perineural invasion and specific tumor location contributed less to the model. We provided an example to illustrate the model's interpretability, presenting the prediction results in an intuitive manner that allows clinicians to clearly observe the weights of the included features in the model's predictions.

In this study, the positive SHAP values indicate that the presence of LVI is associated with an increased likelihood of LNM, highlighting LVI as an important predictor of metastasis risk. The SHAP values for LVI are mainly positive, concentrated on the right side of zero, emphasizing its role in increasing the risk of LNM. Conversely, the lack of negative value distribution further confirms the risk-increasing effect of LVI in predicting LNM. Basic experimental research shows that LVI plays a crucial role in tumor cell invasion. By analyzing lymphatic vessel markers (such as D2-40 and VEGF-C) in tumor tissues, it can be determined that LVI-positive tumors are more prone to LNM [41].

In our study, younger patients were more likely to exhibit LNM. A previous study shows that younger patients have a higher likelihood of LNM due to higher tumor biological activity [42]. This result is consistent with earlier research.

This study demonstrates that tumor size significantly influences the model's prediction of LNM. Larger tumors correspond to positive SHAP values, indicating an increased risk of metastasis, while smaller tumors correspond to negative SHAP values, suggesting a higher probability of no metastasis. Previous studies have shown that larger tumor sizes are generally associated with a higher risk of LNM, whereas smaller tumors have a lower risk. This may be because larger tumors often have deeper invasion depths and higher aggressiveness, increasing the likelihood of cancer cells metastasizing to lymph nodes [43]. Earlier research has indicated that the deeper the invasion depth of cancer cells, the higher the probability of lymph node invasion [31]. The findings of this study align with existing studies, further confirming that tumor invasion depth is positively correlated with LNM risk. Invasion depth is not only an independent predictor of LNM but has also become a key indicator in CRC staging [44].

Additionally, our study reveals a direct relationship between the total number of lymph nodes and patient LNM. The more lymph nodes are removed during radical surgery, the greater the risk of detecting LNM in postoperative pathology. For radical surgery in T1 CRC, previous studies have shown that the number of lymph nodes obtained during systematic lymphadenectomy can be used to assess the comprehensiveness of the surgeon's operation. Removing at least 12 lymph nodes allows for more accurate cancer staging, improved prognostic evaluation, and increased long-term survival rates [45].

4. METHOD

4.1 Data collection and availability

Patients with T1 CRC who underwent radical colorectal tumor resection and lymphadenectomy at a total of seven tertiary medical centers in Anhui and Jiangsu provinces were recruited. 13-year electronic pathology reports were collected

from the First Affiliated Hospital of Anhui Medical University, the First Affiliated Hospital of Bengbu Medical College, Suzhou Municipal Hospital, Jiangsu Cancer Hospital, and the First Affiliated Hospital of the University of Science and Technology of China; over 8-year reports from the Second Affiliated Hospital of Anhui Medical University; and over 6-year reports from the First Affiliated Hospital of Wannan Medical College. These reports had a final diagnosis of CRC.

Each report was carefully reviewed according to the World Health Organization’s definitions, and only those finally diagnosed as T1 CRC with tumor invasion confined to the mucosa and submucosa were included [46]. For the selected T1 CRC cases, we extracted and tabulated demographic information, tumor location, size, gross type, differentiation grade, LVI, perineural invasion, and depth of invasion (Table 2).

Table 2. Summary of patient data

Variables	Classification	All Patients	LNМ (-)	LNМ (+)
N		1205	1082	123
Sex	Male	698	629	69
	Female	507	453	54
Age	<41	31	27	4
	41-60	498	447	51
	>60	676	608	68
Tumor Location	Colon	399	313	86
	Rectum	806	769	97
Tumor Size (cm)	>2 cm	631	572	59
	≤2 cm	574	510	64
	Elevated	1036	941	95
Macroscopic Type	Ulcerating	100	83	17
	Infiltrating	9	8	1
Differentiation	High/Middle	537	494	43
	Middle/Poor	660	583	77
LVI	Present	1108	1034	74
	Absent	97	48	49
Neural Invasion	Present	1192	1075	117
	Absent	13	7	6
Depth of Invasion	Tis+T1a	226	220	6
	T1b	979	862	117

Tumor locations were categorized into the right colon, left colon, transverse colon, entire colon and rectum. The gross types of CRC were classified according to the 2023 CSCO Consensus Guidelines for the Diagnosis of CRC into ulcerative, protruding, and infiltrative types [34]. Additionally, we used the total number of lymph nodes removed during radical colorectal surgery—as recorded in the pathology reports—as a reference index to assess the impact of the surgeon’s proficiency on LNM.

The study protocol was approved by the medical ethics committees of all participating centers and was conducted in accordance with the guidelines of the Declaration of Helsinki. Considering the retrospective observational nature of the study and the anonymity of patient data, written informed consent from patients was not required. We reviewed a total of 10,954 cases, excluding patients with advanced CRC, incomplete clinicopathological data, metastatic CRC, lymphoma, or other life-threatening diseases. Ultimately, 1,205 cases of T1 CRC were selected for inclusion in the study.

4.2 Data preprocessing

All data are statistically categorized. Univariate data with more than two categories are encoded using 0-1 encoding, while multi-category data are encoded using one-hot encoding. All continuous data are standardized.

4.3 Data splitting

K-fold cross-validation (K=10) was used to validate the model’s accuracy. The dataset was randomly divided into 10 mutually exclusive subsets, each containing physiological and

pathological information of 120 patients. In each training round, one subset was selected as the test set, and the remaining nine subsets were used as the training set. The model was trained on the nine training subsets and then evaluated on the test set to assess prediction accuracy. This process was repeated 10 times, recording the performance metrics each time. Finally, we calculated the average of these metrics to represent the model’s stable performance across the entire dataset.

4.4 Hyperparameter tuning

Each model has numerous hyperparameters. Bayesian optimization was employed to ensure the optimal hyperparameter selection for the current model.

4.5 Model selection

Models were trained using all training data and the optimal hyperparameters, with AUC as the primary criterion and also considering the F1 score and accuracy. In this study, LNM (+) indicates the presence of LNM, while LNM (-) indicates its absence. True Positive (TP) represents the number of cases where metastasis was present and correctly predicted; True Negative (TN) represents the number of cases where metastasis was absent and correctly predicted; False Positive (FP) refers to cases without metastasis but incorrectly predicted as having metastasis; False Negative (FN) refers to cases with metastasis that were not identified. Using these values, we calculated accuracy, precision, recall, and F1 score to comprehensively evaluate the model’s classification performance, especially its ability to recognize positive cases

in situations with class imbalance.

4.6 Model interpretation

The SHAP method based on a Nature sub-journal publication [28] was applied for model interpretation. SHAP is a game-theoretic approach to explain machine learning model predictions. Its fundamental principle is to use Shapley values to allocate each feature's contribution to the prediction outcome, providing a unified and fair interpretation framework. Shapley values originate from game theory and are used to fairly distribute the payoff among participants in a cooperative game. They consider each participant's marginal contribution in all possible cooperative combinations.

By calculating the average marginal contribution of each feature across all possible subsets of features, the SHAP value was obtained for each feature. These values represent each feature's contribution to the model's prediction. The sum of all the SHAP values equals the model's output, making the interpretation results intuitive and verifiable.

5. CONCLUSION

This study demonstrates that the CatBoost model can accurately assess the risk of LNM in patients with T1 CRC. Using the SHAP analysis method, it was found that the model's predictions are significantly influenced by features such as LVI, age, total number of lymph nodes retrieved, tumor size, and depth of tumor invasion. In contrast, features like perineural invasion and specific tumor location contribute less to the model. Notably, the analysis indicates that the number of lymph nodes harvested has a crucial impact on LNM; therefore, surgeons should aim to retrieve as many lymph nodes as possible. Combining machine learning with SHAP provides clear and reasonable explanations for personalized risk prediction, enabling clinicians to clearly understand the impact of key features in the model.

REFERENCES

- [1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3): 229-263. <https://doi.org/10.3322/caac.21834>
- [2] Siegel, R.L., Fedewa, S.A., Anderson, W.F., Miller, K.D., Ma, J., Rosenberg, P.S., Jemal, A. (2017). Colorectal cancer incidence patterns in the United States, 1974-2013. *JNCI: Journal of the National Cancer Institute*, 109: djw322. <https://doi.org/10.1093/jnci/djw322>
- [3] Logan, R.F., Patnick, J., Nickerson, C., Coleman, L., Rutter, M.D., von Wagner, C. (2012). Outcomes of the bowel cancer screening programme (BCSP) in England after the first 1 million tests. *Gut*, 61: 1439-1446. <https://doi.org/10.1136/gutjnl-2011-300843>
- [4] Zheng, R., Zhang, S., Zeng, H., Wang, S., Sun, K., Chen, R., Wei, W., He, J. (2022). Cancer incidence and mortality in China, 2016. *Journal of the National Cancer Center*, 2(1): 1-9. <https://doi.org/10.1016/j.jncc.2022.02.002>
- [5] Brierley, J.D., Gospodarowicz, M.K., Wittekind, C. (2017). *TNM Classification of Malignant Tumours*. John Wiley & Sons.
- [6] Nascimbeni, R., Burgart, L.J., Nivatvongs, S., Larson, D.R. (2002). Risk of lymph node metastasis in T1 carcinoma of the colon and rectum. *Diseases of the Colon and Rectum*, 45: 200-206. <https://doi.org/10.1007/s10350-004-6147-7>
- [7] Tomita, N., Ishida, H., Tanakaya, K., Yamaguchi, T., et al. (2021). Japanese Society for Cancer of the Colon and Rectum (JSCCR) guidelines 2020 for the clinical practice of hereditary colorectal cancer. *International Journal of Clinical Oncology*, 26(8): 1353-1419. <https://doi.org/10.1007/s10147-021-01881-4>
- [8] Labianca, R., Nordlinger, B., Beretta, G.D., Mosconi, S., Mandalà, M., Cervantes, A., Arnold, D. (2013). Early colon cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 24: vi64-vi72. <https://doi.org/10.1093/annonc/mdu255>
- [9] Pimentel-Nunes, P., Dinis-Ribeiro, M., Ponchon, T., Repici, A., et al. (2015). Endoscopic submucosal dissection: European Society of Gastrointestinal Endoscopy (ESGE) guideline. *Endoscopy*, 47(9): 829-854. <https://doi.org/10.1055/s-0034-1392882>
- [10] Bosch, S.L., Teerenstra, S., de Wilt, J.H., Cunningham, C., Nagtegaal, I.D. (2013). Predicting lymph node metastasis in pT1 colorectal cancer: A systematic review of risk factors providing rationale for therapy decisions. *Endoscopy*, 45(10): 827-841. <https://doi.org/10.1055/s-0033-1344238>
- [11] Zhu, H.X., Wang, G., Zheng, J.X., Zhu, H., Huang, J., Luo, E.X., Hu, X.S., Wei, Y.J., Wang, C., Xu, A., He, X.Y. (2022). Preoperative prediction for lymph node metastasis in early gastric cancer by interpretable machine learning models: A multicenter study. *Surgery*, 171(6): 1543-1551. <https://doi.org/10.1016/j.surg.2021.12.015>
- [12] Zhang, J.W., Zhang, X.W., Xia, S.J., Dong, Y.J., Zhou, W., Liu, Z.H., Zhang, L., Zhan, W.W., Sun, Y.Z., Zhou, J.Q. (2024). Prediction model for lymph node metastasis in papillary thyroid carcinoma based on electronic medical records. Preprint at Research Square. <https://doi.org/10.21203/rs.3.rs-3909203/v1>
- [13] Kim, B.C., Kim, J.Y., Lim, I., Kim, D.H., Lim, S.M., Woo, S.K. (2021). Machine learning model for lymph node metastasis prediction in breast cancer using random forest algorithm and mitochondrial metabolism hub genes. *Applied Sciences*, 11(7): 2897. <https://doi.org/10.3390/app11072897>
- [14] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6): 386-408. <https://doi.org/10.1037/h0042519>
- [15] Collett, D. (2002). *Modelling Binary Data*. CRC Press.
- [16] Hartwig, M., Bräuner, K.B., Vogelsang, R., Gögenur, I. (2022). Preoperative prediction of lymph node status in patients with colorectal cancer. Developing a predictive model using machine learning. *International Journal of Colorectal Disease*, 37(12): 2517-2524. <https://doi.org/10.1007/s00384-022-04284-7>
- [17] Wang, K., He, H., Lin, Y.Y., Zhang, Y.H., Chen, J.G., Hu, J.C., He, X.S. (2024). A new clinical model for predicting lymph node metastasis in T1 colorectal cancer. *International Journal of Colorectal Disease*,

- 39(1): 46. <https://doi.org/10.1007/s00384-024-04621-y>
- [18] Fujino, S., Miyoshi, N., Kitakaze, M., Yasui, M. (2023). Lymph node metastasis in T1 colorectal cancer: Risk factors and prediction model. *Oncology Letters*, 25(5): 191. <https://doi.org/10.3892/ol.2023.13776>
- [19] Niu, X.Q., Cao, J.Q. (2024). Predicting lymph node metastasis in colorectal cancer patients: Development and validation of a column chart model. *Updates in Surgery*, 76(4): 1301-1310. <https://doi.org/10.1007/s13304-024-01884-6>
- [20] Ichimasa, K., Kudo, S.E., Mori, Y., Misawa, M., et al. (2018). Artificial intelligence may help in predicting the need for additional surgery after endoscopic resection of T1 colorectal cancer. *Endoscopy*, 50(3): 230-240. <https://doi.org/10.1055/s-0043-122385>
- [21] Song, J.H., Kim, E.R., Hong, Y.Y., Sohn, I., Ahn, S., Kim, S.H., Jang, K.T. (2024). Prediction of lymph node metastasis in T1 colorectal cancer using artificial intelligence with hematoxylin and eosin-stained whole-slide-images of endoscopic and surgical resection specimens. *Cancers*, 16(10): 1900. <https://doi.org/10.3390/cancers16101900>
- [22] Takamatsu, M., Yamamoto, N., Kawachi, H., Nakano, K., Saito, S., Fukunaga, Y., Takeuchi, K. (2022). Prediction of lymph node metastasis in early colorectal cancer based on histologic images by artificial intelligence. *Scientific Reports*, 12(1): 2963. <https://doi.org/10.1038/s41598-022-07038-1>
- [23] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H. (2015). Xgboost: Extreme gradient boosting. *R Package Version 0.4-2*, 1(4): 1-4.
- [24] Chen, T.Q., Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Toronto, ON, Canada, pp. 785-794. <https://doi.org/10.1145/2939672.2939785>
- [25] Ahn, J.H., Kwak, M.S., Lee, H.H., Cha, J.M., Shin, H.P., Jeon, J.M., Yoon, J.Y. (2021). Development of a novel prognostic model for predicting lymph node metastasis in early colorectal cancer: Analysis based on the surveillance, epidemiology, and end results database. *Frontiers in Oncology*, 11: 614398. <https://doi.org/10.3389/fonc.2021.614398>
- [26] Piao, Z.H., Ge, R., Lu, L. (2023). An artificial intelligence prediction model outperforms conventional guidelines in predicting lymph node metastasis of T1 colorectal cancer. *Frontiers in Oncology*, 13: 1229998. <https://doi.org/10.3389/fonc.2023.1229998>
- [27] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31.
- [28] Lundberg, S.M., Lee, S.I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- [29] Ji, X.L., Kang, M., Zhao, X.Z., Li, X.Y., Guo, Y.J., Xie, P., Yu, Y.N., Tian, Z.B. (2022). Poorly differentiated cluster grade-A vital predictor for lymph node metastasis and oncological outcomes in patients with T1 colorectal cancer: A retrospective study. *BMC Gastroenterology*, 22(1): 409. <https://doi.org/10.1186/s12876-022-02492-7>
- [30] Bae, H.J., Ju, H., Lee, H.H., Kim, J., Lee, B.I., Lee, S.H., Won, D.D., Lee, Y.S., Lee, I.K., Cho, Y.S. (2023). Long-term outcomes after endoscopic versus surgical resection of T1 colorectal carcinoma. *Surgical Endoscopy*, 37(2): 1231-1241. <https://doi.org/10.1007/s00464-022-09649-1>
- [31] Ueno, H., Mochizuki, H., Hashiguchi, Y., Shimazaki, H., Aida, S., Hase, K., Matsukuma, S., Kanai, T., Kurihara, H., Ozawa, K., Yoshimura, K., Bekku, S. (2004). Risk factors for an adverse outcome in early invasive colorectal carcinoma. *Gastroenterology*, 127(2): 385-394. <https://doi.org/10.1053/j.gastro.2004.04.022>
- [32] Miyachi, H., Kudo, S.E., Ichimasa, K., Hisayuki, T., et al. (2016). Management of T1 colorectal cancers after endoscopic treatment based on the risk stratification of lymph node metastasis. *Journal of Gastroenterology and Hepatology*, 31(6): 1126-1132. <https://doi.org/10.1111/jgh.13257>
- [33] Ikematsu, H., Yoda, Y., Matsuda, T., Yamaguchi, Y., et al. (2013). Long-term outcomes after resection for submucosal invasive colorectal cancers. *Gastroenterology*, 144(3): 551-559. <https://doi.org/10.1053/j.gastro.2012.12.003>
- [34] National Health Commission of the People's Republic of China. (2023). Chinese protocol of diagnosis and treatment of colorectal cancer (2023 edition). *Chinese Journal of Surgery*, 61(8): 617-644. <https://doi.org/10.3760/cma.j.cn112139-20230603-00222>
- [35] Washington, M.K., Berlin, J., Branton, P., Burgart, L.J., Carter, D.K., Fitzgibbons, P.L., Halling, K., Frankel, W., Jessup, J., Kakar, S., Minsky, B., Nakhleh, R., Compton, C.C. (2009). Protocol for the examination of specimens from patients with primary carcinoma of the colon and rectum. *Archives of Pathology and Laboratory Medicine*, 133(10): 1539-1551. <https://doi.org/10.5858/133.10.1539>
- [36] Zlobec, I., Lugli, A. (2018). Tumour budding in colorectal cancer: Molecular rationale for clinical translation. *Nature Reviews Cancer*, 18(4): 203-204. <https://doi.org/10.1038/nrc.2018.1>
- [37] Nagtegaal, I.D., Quirke, P. (2008). What is the role for the circumferential margin in the modern treatment of rectal cancer? *Journal of Clinical Oncology*, 26(2): 303-312. <https://doi.org/10.1200/JCO.2007.12.7027>
- [38] Yoshii, S., Nojima, M., Noshio, K., Omori, S., Kusumi, T., Okuda, H., Tsukagoshi, H., Fujita, M., Yamamoto, H., Hosokawa, M. (2014). Factors associated with risk for colorectal cancer recurrence after endoscopic resection of T1 tumors. *Clinical Gastroenterology and Hepatology*, 12(2): 292-302. <https://doi.org/10.1016/j.cgh.2013.08.008>
- [39] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5): 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- [40] Molnar, C. (2020). *Interpretable Machine Learning*. Leanpub. https://originalstatic.aminer.cn/misc/pdf/Molnar-interpretable-machine-learning_compressed.pdf
- [41] Zhang, L.J., Deng, Y.X., Liu, S.R., Zhang, W.L., Hong, Z.G., Lu, Z.H., Pan, Z.Z., Wu, V.J., Peng, J.H. (2023). Lymphovascular invasion represents a superior prognostic and predictive pathological factor of the duration of adjuvant chemotherapy for stage III colon cancer patients. *BMC Cancer*, 23(1): 3. <https://doi.org/10.1186/s12885-022-10416-7>

- [42] Xie, X., Yin, J., Zhou, Z., Dang, C., Zhang, H., Zhang, Y. (2019). Young age increases the risk for lymph node metastasis in patients with early colon cancer. *BMC Cancer*, 19: 803. <https://doi.org/10.1186/s12885-019-5995-4>
- [43] Hu, S.D., Li, S.Y., Teng, D., Yan, Y., Lin, H.G., Liu, B.Y., Gao, Z.H., Zhu, S.Y., Wang, Y.F., Du, X.H. (2021). Analysis of risk factors and prognosis of 253 lymph node metastasis in colorectal cancer patients. *BMC Surgery*, 21: 280. <https://doi.org/10.1186/s12893-021-01276-2>
- [44] Amin, M.B., Edge, S.B., Greene, F.L., Byrd, D.R., et al. (2017). The eighth edition AJCC cancer staging manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA: A Cancer Journal for Clinicians*, 67(2): 93-99. <https://doi.org/10.3322/caac.21388>
- [45] Hashiguchi, Y., Hase, K., Ueno, H., Mochizuki, H., Shinto, E., Yamamoto, J. (2011). Optimal margins and lymphadenectomy in colonic cancer surgery. *Journal of British Surgery*, 98(8): 1171-1178. <https://doi.org/10.1002/bjs.7518>
- [46] Nagtegaal, I.D., Odze, R.D., Klimstra, D., Paradis, V., Rugge, M., Schirmacher, P., Washington, K.M., Carneiro, F., Cree, I.A. (2020). The 2019 WHO classification of tumours of the digestive system. *Histopathology*, 76(2): 182-188. <https://doi.org/10.1111/his.13975>